DOCUMENT RESUME

ED 244 980 TM 840 289

AUTHOR Gipps, Caroline; Goldstein, Harvey
TITLE Local and National Testing in the UK: The Last Ten Years.
PUB DATE Apr 84
NOTE 24p.; Paper presented at the Annual Meeting of the American Educational Research Association (68th, New Orleans, LA, April 23-27, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Standards; College Entrance Examinations; Educational Assessment; *Educational Testing; *Educational Trends; Foreign Countries; *Handicap Identification; Instructional Improvement; *National Programs; *School Districts; Screening Tests; Testing Problems; Testing Programs; Test Interpretation
IDENTIFIERS *Assessment of Performance Unit (United Kingdom); *United Kingdom

ABSTRACT

New developments in testing in the United Kingdom (UK) since 1965 are described. Standardized testing at the local level declined dramatically with the widespread introduction of comprehensive secondary education. However, in the late 1970's widespread local testing programs were re-introduced for the purposes of monitoring student progress, screening students to identify those in need of special help, or providing information for transfer from junior to senior school. A national testing program, the Assessment of Performance Unit (APU), was established in 1974. It is designed to assess achievement in language, math, science, and modern language. The emphasis of the APU has shifted away from its original purpose of providing information relevant to policy making and resource allocation toward providing detailed information to guide teaching practice. In the UK, there are also two types of public examination: the General Certificate of Education Ordinary Level (at age 16) and Advanced Level (at 18), and the Certificate of Secondary Education (at 16) for the less academic student. These examinations are set by various examination boards, and with such a diverse system, there are questions over comparability and confusion over whether the grades awarded are norm-referenced or criterion-referenced. (BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED244980

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

LOCAL AND NATIONAL TESTING IN THE UK:

THE LAST TEN YEARS

Caroline Gipps and Harvey Goldstein

University of London Institute of Education

London U.K.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
C. V. Gipps
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented to the American Educational Research Association

Annual Meeting, New Orleans, April 1984.

Tm 840 283

Author Address:- Dr Caroline Gipps & Professor Harvey Goldstein
Institute of Education
18 Woburn Square
London WC1H ONS
United Kingdom

Local and National Testing in the UK: The Last Ten Years

Caroline Gipps and Harvey Goldstein

Local Testing Programmes

In the period after 1965 standardised testing at local level in this country declined dramatically following Government Circular 10/65, the document which heralded the widespread introduction of comprehensive secondary education. Previously the 11+ test, a series of IQ, English and Maths tests, had been used to allocate children to academic (grammar) or non-academic (secondary modern) secondary schools. As the process of comprehensivisation proceeded the need for selection declined. However, some school districts (local education authorities - LEAs) continued with attainment and ability testing in order to ensure a 'balanced' intake into secondary schools, for example the Inner London Education Authority. Other LEAs never entirely abandoned selection, retaining one or two grammar schools for which entry was determined through testing and a handful retained a completely selective system. Thus group testing did not disappear altogether but in the mid to late 70s several events, social, political and educational, led to an increase in concern over education which culminated in the introduction of widespread testing programmes around 1978. The significant events were - local authority reorganisation in 1974, the Bullock Report in 1975, the 'Black Papers' in 1975 and 1977, the Prime Minister's Ruskin College speech in 1976, and the William Tyndale school report in 1976. We shall deal with each of these events in turn.

Following local authority reorganisation new school districts were created and as a result administrators and professionals in areas which had taken on new schools felt that they did not have a grasp of the levels

of attainment of the new children and schools in their area. This was compounded by the fact that with the ending of the 11+ there was no information on levels of performance of children leaving primary (elementary school from 1st through 7th grade) school. The reduction of 11+ testing had, in fact, had considerable impact at the primary school level. Freed from the constraints of this leaving exam, there was a revolution in the primary school curriculum with child-centred approaches, discovery learning and an emphasis on individual or small group teaching. The reduced emphasis on traditional methods of teaching the basic skills had its critics and the Black Papers edited by Dr Rhodes Boyson (now Secretary of State for Social Services, previously a minister of Education) argued that modern methods in the primary school, and non-selective secondary education, were resulting in a lowering of 'standards'. The Bullock Committee on the teaching of English was set up because of concern over standards of literacy, and one of the Committee's recommendations was that LEAs should monitor reading levels regularly through the use of standardised group reading tests. This report, produced by a committee of highly regarded professionals, became one of the most influential documents of the 70s and effectively it set the seal of approval on testing by LEAs. This was just as well, for with the increased calls for information on standards by politicians at this time came an event which was known as the 'William Tyndale affair'. A primary school in London which ran a progressive regime became the focus of concern for some of the parents and staff. There was an enquiry into the running of the school and the Head and some of its staff were suspended from duty; this was an unheard-of happening and the shock waves ran through the education system. "Could it happen here?" asked many a Director of Education (Superintendent of Schools) and if they did not know what standards in the basic skills were in their primary schools they set about designing monitoring systems. The simplest monitoring scheme of course is a group testing programme, and as the Bullock Committee

had advocated regular monitoring of reading it was not too hard to get this idea across to schools. Then in 1976 the Prime Minister James Callaghan made a speech at Ruskin College in Oxford in which he questioned the right of educationists to determine the direction in which education was going without reference to other interested parties. This was the start of the accountability debate in this country which was related to issues of value for money in the new era of public expenditure constraints: "We spend £6 billion a year on education so there will be discussion" (TES 22.10.76) said Prime Minister Callaghan.

Not all these precipitating factors had the same impact in different LEAs. Some LEAs cite political factors for setting up testing programmes - the atmosphere in the mid to late 70s at the time of the Ruskin College speech, the Black Papers and the William Tyndale affair resulting in pressure from local politicians; others cite organisational factors - the ending of the 11+, secondary school reorganisation and LEA reorganisation all leading to a demand for information particularly relating to primary/secondary transfer; yet others cite professional factors - concern over the number of children referred for remedial help, both too large and too small, and concern over reading standards following publication of the Bullock Report. But whatever the initiating factors, by 1981 almost 80% of all LEAs had a testing programme* (79 out of 104), with 1978 being the year when most new testing programmes were introduced (Gipps et al, 1983). Briefly, the situation with regard to who is tested on what is similar to the USA (Wigdor and Garner, 1982) with most testing taking place at junior (4th through 7th grades) level, most tests covering the basic skills in reading, maths and language, and norm-referenced tests being more popular than criterion-referenced tests.

---

* defined as .. "any testing of children in an age group which is organised and promoted by the LEA as a matter of policy".

Regardless of reasons for introduction, most testing programmes are ascribed a variety of purposes: monitoring - that is of overall standards within an area and/or of individual schools; screening - that is to identify children who are in need of special help or provision; and providing information for transfer from junior to senior school, are the most widely given reasons for testing, and not individually either. We were struck by the range of reasons given for testing: LEAs seem to believe that testing, and nearly always a single test, can satisfy several purposes. We received comments such as "The tests are administered to monitor performance across the authority, to indicate resource requirements and to enable decisions to be made on appropriate curriculum". All this on the basis of scores on a simple non-diagnostic reading test. Our findings indicated a lack of clear thinking in LEAs as to why they had their testing programmes, which is a disquieting fact in itself but also there are technical limitations to the efficiency with which the same test can be used simultaneously to monitor and to screen. Fresh thinking about testing has perhaps been hard for LEAs because the testing of reading has for so long been part and parcel of schooling in the UK. What seems to have happened in some authorities is that programmes which set out originally with screening as their main purpose have, as often as not, had monitoring added, perhaps as a political response and then, with the threat of cuts, had allocation of resources added as well. The implementation of the 1981 Education Act in 1983 which obliges LEAs to ensure that all children with special educational needs are catered for adequately, preferably within mainstream schools, emphasises the identification and assessment of children with special educational needs, so we may see an expansion of testing for screening purposes.

Last year (1983) we carried out a survey of screening programmes in LEAs and found that around 70% (72 out of 104 LEAs) did indeed use

standardised tests to identify children with special educational needs (and a further three LEAs are planning to introduce test for screening/ identification purposes). These screening programmes are little different in outward appearance from the accountability testing programmes we were told about in 1981 and we have yet to find out whether the results are used in any more rigorous a fashion. We found then that results of monitoring programmes were not used to make schools or teachers accountable in any hard-line way. There was no LEA which published 'league-tables', i.e. named school results (though one LEA attempted to do this recently, TES 2.12.83). The chief reason for this was that teachers, individually, on testing panels convened by LEAs, and through their Unions made it abundantly clear that league tables were not acceptable, and LEA officers in turn did much to persuade Education Committee members (i.e. representatives of the community and local politicians) not to ask for league tables. At the school level we found, as did Leslie Salmon-Cox and colleagues in Pittsburg (1981) and Keilaghan and colleagues in Ireland (1982), that teachers made little direct use of test scores themselves. Scores were put into record books largely for the benefit of someone else, though of course if the scores gave cause for concern teachers would act on them, but by and large the feeling was that they were of use mostly to someone else - the Head or the LEA. The need of both heads and local authorities to have testing in order to keep a check on 'standards' (whatever that means) was well accepted in the school system. In that situation the perceived need is for norm-referenced tests for the comparisons that they make possible. One is tempted to say that, given no-one expects results from this type of testing to be of much use except for level-checking and comparison purposes, the need is for tests which are as quick, simple and straightforward as possible with perhaps less concern for maintaining reliability and validity. We are not of course here talking about individual diagnostic testing such as that carried out by educational psychologists or special education staff.

The mood now at local level has moved away from concerns with monitoring and accountability to - in the face of expenditure cuts and falling rolls, resource allocation (though it is far from clear how testing information can help in this area) and now into the special educational needs area.

Special Educational Needs

Our own interests have also turned towards children with special educational needs and methods of identification of such children. The official definition is of little help here*: 'a child has special educational needs if he (sic) has a learning difficulty which calls for special educational provision to be made for him ...' and 'a child has a "learning difficulty" if ... he has a significantly greater difficulty in learning than the majority of children of his age' (DES 1981).

As we have already said, many LEAs have fallen back on the good old standardised reading test as a first line of attack in their attempts to fulfil their now wide obligations to provide adequate support for all children with special educational needs. With the norm-referenced overtones of the official definition this may be appropriate but there is also a move towards using skills or curriculum-based assessments with precisely stated objectives, on the basis that clear objectives combined with feedback on progress are a necessary prerequisite for effective teaching (Cameron, 1982). This movement stems partly from the objectives approach of much special education teaching, partly from a genuine desire to develop assessment techniques which provide some feedback for the teacher and which she/he can actually use, and partly from changing models of provision for children

---

* It is an interesting aside here that Sir Cyril Burt, that much denigrated English psychologist, also had trouble with vague definitions as far back as 1921. In those days the statutory definition was 'incapable of receiving proper benefit from instruction in the ordinary public elementary schools' (Burt 1921 p167). See Gipps & Goldstein (1984) for expansion of this theme.

with special educational needs. The current wisdom is that, in the face of doubts about the effectiveness of remedial teaching based on withdrawal sessions by peripatetic staff, children with learning difficulties are best helped by their regular teacher in the context of their own classroom. The catch-phrase is 'All teachers are teachers of children with special needs'. In this situation there is a need to provide the class teacher with assessment materials that are curriculum-based and therefore help the teacher to design and implement teaching programmes matched to each child's special needs (Ainscow & Muncey, 1983). This seems a potentially interesting development in what we might call 'useful' testing (as opposed to use-for-others testing which we described earlier) and our current research is involved in investigating such new developments.

These developments are in line with the recommendations about assessment in the 1982 NAS panel report on Mild Mental Retardation classification/ placement.

> "The fundamental assessment principal emphasized repeatedly ... was educational utility. Information related to educational decision making, especially that which leads to more effective educational programming, was seen as worthwhile, beneficial ... Messick's well-placed emphasis on assessing the regular education program before or concurrent with initial referral as well as development of interventions in regular education as a first step is in line with current legal, legislative and professional opinion. Moreover, fiscal realities, in addition to perceptions of children's best interests, dictate greater use of interventions within regular education instead of referring all (or even most) problems to very expensive special education programs."
>
> (Reschly, 1983)

Many of the issues raised by Lorrie Shephard in her NCME presidential address last year are relevant here too; for example, the technical inadequacy of tests used in assessment, the professionals' poor awareness of the difference between an adequate and inadequate test, traditional test choice preferences in the face of evidence of inadequacy (see particularly Steadman & Gipps, 1984) and the widely felt need for norms (Shephard, 1983).

National Monitoring

In England the Assessment of Performance Unit (APU) parallels the American National Assessment of Educational Progress. The APU was set up by the Department of Education and Science (DES) in 1974, but there had been a considerable gestation period and its appearance was the result of many of the same concerns which caused LEAs to set up their own monitoring systems.

From 1948 to 1964 the DES had commissioned regular national reading surveys to be carried out by the National Foundation for Educational Research. These showed that during the 16 years of the surveys there had been advances of several months in the reading ages of 11 and 15 year olds. The next survey was not conducted until 1970 and, unlike the previous surveys, it did not show an advance in average reading ages. That survey was bedevilled with problems which resulted in a sample which was probably unrepresentative, and the test which had been used in previous surveys was by then out of date. Nonetheless, these results caused a furore in the world of education and beyond, and critics of progressive education took it as evidence of a deteriorating system of state schooling. One of the consequences was the setting up of the Bullock Committee, to which we have already referred. and the report's first recommendation was

for a system of national monitoring employing new instruments. This was the first public indication that central government was interested in national monitoring, but in fact discussions had been going on behind the scenes for some time. As early as 1968 an internal DES paper suggested a wide ranging testing programme as one means of assessing the results of educational investment, and maths was singled out as an area in which to start (DES, 1971). As we have seen, these discussions took place against a backdrop of increasing concern over standards. In the face of this the DES' lack of control over what went on in schools, in spite of the fact that the DES funded schools and was ultimately accountable for what went on in them, caused increasing concern amongst some officials. A national scheme of monitoring would provide the DES with some means of evaluating the performance of the education system directly (so the reasoning went) and hence possibly with an indirect say in curriculum content. It might also provide longed-for evidence to dispute the claims of those who argued that standards were falling (Gipps & Goldstein, 1983).

However, the APU was actually announced in the context of government moves to deal with educational disadvantage and the educational needs of immigrants (DES, 1974) and the APU's role was to help to develop criteria to identify educational disadvantage. This announcement caused few ripples at the time since the move to deal with disadvantage and underachievement was welcomed by educationists. The early publicity material put out by the APU, however, had a different tale to tell: the APU's role was to monitor in order to provide information on standards and how these change over time. The educational climate in the mid 70s was, as we have seen, one in which the professionals were being criticised, at least indirectly. In this climate the APU became the focus of wider attention and suspicion on behalf of many of those in education, linked as it was inevitably with the move towards greater accountability. Proposals to monitor standards

nationally were perceived to emanate from the political right and were threatening to the teaching profession.

There were two main areas of concern. The first was that, though ostensibly concerned with children's standards, the APU was really dealing with teachers' competence. The second concern was its possible effect on the curriculum, through the curriculum models adopted by the testers, the increased importance of the areas tested (and by corollary the decreased importance of areas not tested), and teaching to the test.

Before looking at what came of these concerns, a brief look at what the APU is and how it works. The APU is in fact a small unit within the DES. It oversees the surveying of performance, which is actually contracted out to the NFER, Leeds University and Chelsea College (University of London). Each of the test development teams has an advisory group, there is a Statistical Advisory Group which advises the Unit on technical matters, and a Consultative Committee which is largely made up of non-DES people and is representative of outside interests. This latter committee makes suggestions about policy matters and has been extremely influential.

The APU does not test in as many areas as the NAEP; it covers language, maths, science and modern language, essentially a core curriculum, but it may come to include design and technology. It tests only at three ages 11, 13 and 15 (and not at all ages in all subjects). Initially maths, language and science monitoring was carried out annually for five years and modern language for three years. This initial cycle of five annual surveys ended in 1982 for maths, 1983 for language and 1984 for science; the last of the three annual surveys in modern language will take place in 1985. Maths, language and science will then be monitored every five years; a decision about the future of modern language monitoring has yet

to be made. This rolling system of monitoring with maths, language and science taking it in turns will have the function of updating the national picture and identifying trends, while limiting the burden on schools and reducing costs.

The take-up of published reports, particularly by teachers, has been poor and this was a major theme or our 1982 evaluation (Gipps & Goldstein, 1983, op cit). In November 1982 the decision was made by the DES that more emphasis be put on dissemination. Thus there is now a series of occasional papers and a regular newsletter, rather in the style of the NAEP newsletter. The DES has produced a booklet on the writing performance of 15 year olds and the Association for Science Education, acting as the APU's agent, is publishing a series of pamphlets on science performance aimed at the classroom teacher.
Instead of publishing major reports at considerable expense, the emphasis is now on short, easy to read booklets on specific areas aimed at a specific audience. The APU has also commissioned independent evaluations of the maths and language reports following the NAEP model. It is expected that these will result in various documents for in-service training.

There has been continuing discussion within the Unit, its committees, groups and teams over the nature and extent of the background variables which should be measured. Information of this sort is essential for the interpretation of findings and to provide data of value to policy makers, which is part of its task, that is, to identify differences in achievement in relation to the circumstances in which children learn. The Statistics Advisory Group has advised against the collection of several proposed variables because of problems of measurement, while the Consultative Committee has been consistently against the collection of home background information from either parents or children. The current situation is that school-based background measures are being collected by the teams in their

surveys. However, composite measures of background (social and educational) have limited potential for explaining performance at an individual level. And, as Nuttall (1983) has pointed out, there can be little doubt that in any case information on classroom processes and detailed curriculum information is vital for interpretation of survey results. Such data is not easy to obtain from large scale surveys but requires more intensive in-depth studies. In the fallow four year period between surveys, the teams will now have an opportunity to make in-depth studies, which were promised when the work was first commissioned. At this point, however, only in-depth analysis of existing data is involved; although in-depth studies involving the collection of new data are possible this option has not yet been taken up by the test development teams.

This problem is not restricted to the UK national assessment programme. A comparison of the American, British and Australian monitoring programmes by Power and Wood (1984, in press) concluded:

> "There is no way in which a national assessment program of the type developed could serve a social accountability function, given the structure and politics of education in Australia, the UK and the US. As well, in developing the programs political considerations proved more important than clarification of objectives and of what would be needed if these were to be met. As a consequence, the programs developed into bland monitoring exercises of little direct information value to policy makers and educators.
>
> ... As all three evaluations suggest, the picture would be clearer and more readily interpretable if additional student and school background and process data were collected and further research on the instruments and follow-up studies were undertaken ..."

The other area in which the APU (and NAEP) has met problems is in analysing and reporting changes in performance over time. There is no consensus on how to analyse trends over time and this relates directly to the issue of what one can say about standards (in terms of whether they are rising or falling which is what most people want to know). As Nuttall (op cit) says "Finally, the measurement of change of over time: the only possible conclusion is that a satisfactory long-term method has not been devised". NAEP, for example, has relied on using a number of items that are common from one survey to the next to indicate change, although ETS seems to be proposing to resurrect latent trait models for this purpose - a proposal contemplated but now rejected by the APU. The main problem with using a common core of items is that this method cannot provide a wholly representative sample of the items used in any particular survey and so the information thus provided on changes in performance over time is inevitably limited. The APU teams are also using some common items from one survey to another, for example, in maths half the items were common in the first and last annual surveys. At the end of the five year period of surveying, each team will produce composite measures of performance over the five years which will serve as a baseline (or standard) with which to compare performance measured subsequently in the five-yearly surveys. By then, the question of how to analyse trends in performance may have been answered in part. Certainly the Unit, although it said much about standards in the early days, has not attempted to define 'standards' in the sense of acceptable or looked-for performance and will instead rely on describing measured performance over a period of several years, a far less contentious and more acceptable task, and on comparing relative changes between groups, e.g. sexes, over time (Goldstein, 1983). The DES however, is not quite so circumspect: the pamphlet on the writing performance of 15 year olds was launched as a

contribution to the debate on standards "to trigger a public debate about the content of English teaching and the standards needed " (TES, 2.12.83, p3).

By adhering to the principle of light sampling, anonymity of students and schools, and the inclusion of teacher union representatives on its Consultative Committee, the APU has gone a long way towards allaying teachers' early fears. The extent to which the APU has carried the teachers with it can be illustrated by some findings of a teacher-interview survey we carried out in late 1982: approximately 70% of the primary and secondary heads interviewed (120) were in favour of national monitoring (Gipps et al, 1983, op cit) with accountability and the need to keep a check on standards to the fore in their comments.

The other early concern was about its impact on the curriculum, specifically its role in introducing a core curriculum and the curriculum backwash effect of the test items used. The APU's sampling and testing policies have prevented the curriculum backwash which would result from teaching to the test. The impact of the APU on the development of a core curriculum, however, cannot easily be separated from the influence of other factors in education. In 1982, when we wrote our evaluation of the APU, we felt that any impact there might be on the curriculum would be via the curriculum models adopted by the test development teams; the teams were aware of this and operated on a wide curriculum model so that any impact would be widening and not narrowing, and positive not negative. Indeed in 1982 there was a certain ambivalence on the part of the APU towards its role vis à vis the curriculum. The APU had been accused of being a Trojan horse to bring in an assessment-led curriculum; this however was a slightly paranoid view of the role of central government in the education system without sufficient awareness of the constraints on

it through the countervailing power of bodies such as the National Union of Teachers. But the Unit, in order to allay fears, maintained that it would not attempt to influence the curriculum via backdoor methods. That ambivalence about its current role has now gone and one of the Unit's current major aims is to milk its very detailed survey findings in order to improve curriculum content and delivery ,that is, teaching. It hopes to achieve this via its new dissemination policy and by running in-service courses for teachers and LEA subject advisers.

Indeed, the whole curriculum scene has changed over the last two years, since the Schools Council, the teachers' body responsible for development of the curriculum and examinations, has been disbanded and two new organisations have been set up - the School Curriculum Development Committee and the Secondary Examinations Council - with more DES control. Though there are no formal links between the APU and these two organisations the APU data will be fed into their committees to help them in their early deliberations. Two particular areas of input are likely to be in helping to think about criteria for allocating grades in the new 16+ exams and in suggesting modes for examining. Of course, now the DES has the SCDC, it no longer needs the APU as a means of having some say in the curriculum.

Within the Unit the emphasis now seems to have shifted away from a concern with information relevant to policy making and resource allocation. Instead it is in providing detailed information to guide teaching practice that the APU's profile seems to be highest. The incidence of low achievement, changes over time, policy decisions concerning resource allocation, making test items available to LEAs these are all still on the agenda but one senses that they are no longer considered to be paramount. These areas are of course potentially far more problematic, particularly given the way the APU carried out its tasks prior to 1982.

Current APU moves to disseminate its findings to improve the curriculum - by, it must be admitted, anything but backdoor methods - can be given a cautious welcome (and certainly the demand from LEAs and teachers for courses and conferences seems quite considerable). However, its future impact on the curriculum is uncertain and much will depend on the APU's links with the aforementioned new organisations - the SCDC and the SEC - and how these attempt to shape the curriculum.*

New Developments in Public/School-leaving Examinations

So far we have not mentioned the area of public examinations - those which students take at 16 and 18. There are two types of exam, the General Certificate of Education (GCE) Ordinary Level (at 16) and Advanced Level (at 18) which are meant for the top 20% of the population. For the less academic student there is the Certificate of Secondary Education (CSE) taken at 16 only. These exams are set by various examinations boards, independent bodies under the aegis of the Universities, except for one type of CSE exam 'Mode III' which is set by the student's own school but has to be approved by the relevant examinations board. With such a diverse system, there are bound to be questions over comparability and there is confusion over whether the grades awarded are norm-referenced or criterion based. In fact they are largely norm-based, i.e. the top x% always get Grade A, although some variation is allowed

---

* Wood and Power (forthcoming) make the point that indeed, given the human and financial resources the APU has received, it is not surprising the APU has been able to produce superior test materials. Now that big curriculum reform projects have gone out of favour in the UK, the APU can be viewed as the "nearest thing to a curriculum reform project".

from year to year so as not to penalise an apparently unusually highly-scoring group, and there have been distinct trends over the years in some subjects.

Although the top grade in CSE has the same value as a pass grade in GCE 'O' level, the latter has become the qualification looked for by employers and the CSE has as a result become devalued. There has in fact been considerable dissatisfaction with this dual system of examining and in 1976 the Schools Council submitted proposals to develop a single examination at 16+. In 1983 the Government agreed to introduce such a single system subject to the creation of satisfactory national criteria for syllabuses, assessment procedures and the award of grades. A major exercise to develop such criteria is underway on the part of the exam boards, the SEC and other national bodies (Orr and Nuttall,1983).

The present Secretary of State for Education has brought in the notion of grade-related criteria: "national criteria must be established ... to ensure that ... all boards apply the same performance standards to the award of grades" (Orr and Nuttall, op cit). This development, which has not yet been completed is part of a more general trend to move away from purely norm-referenced testing towards criterion-referenced testing - which attempts to specify more precisely what a student can actually do. The attraction of criterion-referenced testing is that it can have a positive value for all students, since it is a record of what can actually be done. Nevertheless, the practical pressures to aggregate a large number of criterion-referenced assessments for purposes of selection and so on, is likely to leave us with many problems - not the least of which is the requirement for comparability. Indeed, the distinction between norm-referenced and criterion-referenced testing in widely misunderstood in

the UK (see Black et al, 1984) and it seems likely that the present high level political advocacy of criterion-referenced testing and its acceptance by much of the teaching profession is based on a misunderstanding of its nature and potentialities (Goldstein, 1984).

One particular type of criterion-referenced test has existed in this country for many years - the graded test. The most well known example of graded testing is that of music, though there are now moves underway to develop graded tests in other areas, for example, foreign languages (which has actually been going on for some time) and English. It is likely, however, that subjects like maths lend themselves more readily to graded testing (with pre-specified criteria) than subjects like English (Nuttall and Goldstein, forthcoming).

Another approach is that of profiling in which an individual's results in a subject are reported in the form of a profile which specifies levels of attainment in each/range of skills (see Mortimore and Mortimore, 1984, for a review of profiling and graded tests). However, this approach is not without its measurement problems either; as Nuttall and Goldstein (op cit) point out, one of the more serious of these is how to deal with aggregating very detailed assessments of individual attributes. At another level, Her Majesty's Inspectors are concerned that schools will seize on profiles and use them without careful planning, not to mention dealing with the issue of comparability between schools (Education 29.7.83).

There is also the danger that profiles and graded tests will be used only for the bottom 40% of the ability range. Indeed, at the end of 1982 the Government made available £½ million for the development of graded tests in maths for lower attaining pupils (DES Press Notice 268/82). One is here driven to question the motives of a Government which is

encouraging the development of a unified examination system on the one hand and development of graded tests for a particular section of the population on the other.

developments
One thing all these new / have in common is the desire to move towards an examining system which tells us something about what students can do in specific terms. This parallels the move in special/remedial education towards curriculum- and teaching objectives-based assessment. The underlying requirement is that test scores and exam results should carry more information with them than they do at present. It would be by no means a bad thing if these scores and results were to be more useful, and therefore used more, than scores from norm-referenced standardised tests. One of the challenges to those concerned with educational measurement is in finding ways in which such sets of more detailed information can be conveyed informatively.

In January of this year the Education Secretary made a major speech on future educational policy which received warm welcome from many in education. This significant speech, known as the 'Sheffield speech', emphasised the need to raise standards and outlined the changes required in examining and the curriculum in order to achieve this rise. The Secretary of State gave as his objective bringing 80-90% of all 16 year old pupils at least (his emphasis) up to the level now associated with that grade in CSE which is currently achieved by average pupils. He reiterated his call for a greater degree of criterion-referencing in public exams, and for explicit definitions of the objectives of each phase, and of each subject area, of the curriculum. Explicitly defined curricular objectives increase teacher expectations, so his argument went, and high expectations based on defined objectives motivate pupils to give of their best. And - echoes of the then Prime Minister's Ruskin College

speech in 1976 - "There would be a further gain if defined curricular objectives were not only broadly agreed by all the partners in the education service but were also shared by those who use it and pay for it - parents, employers, and the tax and ratepaying public" (our emphasis). Although the emphasis on standards and value for money is much the same as it was in 1976, something has changed: the current view about who has a right to be involved in the curriculum. "There is now no serious dispute that the school curriculum is a proper concern not only of the teachers, but also of parents, governing bodies, LEAs and the Government .." (Education 13.1.84).

Testing, and its inevitable companion the curriculum, has come a long way in the last ten years.

REFERENCES

AINSCOW M & MUNCEY J (1983) All Teachers are Teachers of Children with Special Needs Elm Bank Teachers' Centre, Coventry LEA.

BLACK P, HARLEN W & ORGEE T (1984) Standards of Performance - Expectations and Reality APU Occasional Paper No 3, London: DES.

BURT C (1921) Mental and Scholastic Tests London: P S King & Sons.

CAMERON R J (1982) Teaching and Evaluating Curriculum Objectives, Remedial Education 17 pp 102-108.

DES (1971) Report of the Working Group on the Measurement of Educational Attainment.

DES (1974) Educational Disadvantage and the Educational Needs of Immigrants Cmnd 5720 London: HMSO.

DES (1981) Education Act 1981 London: HMSO.

GIPPS C & GOLDSTEIN H (1983) Monitoring Children: An Evaluation of the Assessment of Performance Unit London: Heinemann Educational Books.

GIPPS C & GOLDSTEIN H (1984) Twenty per cent with special needs: another legacy from Cyril Burt? (in press)

GIPPS C, STEADMAN S, BLACKSTONE T & STIERER B (1983) Testing Children: Standardised Testing in Local Education Authorities and Schools, London: Heinemann Educational Books.

GOLDSTEIN H (1983) Measuring Changes in Educational Attainment over Time: Problems and Possibilities, J Educational Measurement, 20, 4.

GOLDSTEIN H (1984) Models for Equating Test Scores and for Studying the Comparability of Public Exams, in Assessing Educational Achievement, Nuttall, D.(Ed), Falmer Press (in press).

KELLAGHAN T, MADAUS G & AIRASIAN P (1982) The Effects of Standardised Testing, Boston: Kluwer-Nijhoff Publishing

MORTIMORE J & MORTIMORE P (1988) Secondary School Examinations:'the helpful servants, not the dominating master' Bedford Way Paper No 18, London: University of London Institute of Education.

NUTTALL D L (1983) Monitoring in North America Westminster Studies in Education Vol 6, 1983.

NUTTALL D L & GOLDSTEIN H (1984) Profiles and Graded Tests: the technical issues, in Profiles in Action, London: Further Education Unit (forthcoming)

ORR L & NUTTALL D L (1983) Determining standards in the proposed single system of examining at 16+ Comparability in Examinations Occasional Paper 2, London: Schools Council.

POWER C & WOOD R (1984) National Assessment: A Review of Programs in Australia, United Kingdom and United States, Comparative Education Review (in press).

REFERENCES (continued)

RESCHLY D J (1983) Comments on the National Academy of Sciences Report on Mild Mental Retardation Classification/Placement, paper presented to AERA Annual Meeting, Montreal, 1983.

SALMON-COX L (1981) Teachers and standardised achievement tests: what's really happening?, Phi Delta Kappan, May 1981.

SHEPHARD L (1983) The Role of Measurement in Educational Policy: Lessons from the Identification of Learning Difficulties, Educational Measurement: Issues and Practice, Fall 1983.

STEADMAN S & GIPPS C (1984) Teachers and Testing: pluses and minuses, Educational Research (in press).

WIGDOR A & GARNER W (Eds) (1982) Ability testing: uses, consequences and controversies, Washington DC: National Academy Press.

WOOD R & POWER C (forthcoming) National Assessments and 'standards': for better or worse?